

## Article

# Exploration of the Solubility Hyperspace of Selected Active Pharmaceutical Ingredients in Choline- and Betaine-Based Deep Eutectic Solvents: Machine Learning Modeling and Experimental Validation

Piotr Cysewski , Tomasz Jeliński  and Maciej Przybyłek 

Department of Physical Chemistry, Pharmacy Faculty, Collegium Medicum of Bydgoszcz, Nicolaus Copernicus University in Toruń, Kurpińskiego 5, 85-096 Bydgoszcz, Poland; tomasz.jelinski@cm.umk.pl (T.J.); m.przybylek@cm.umk.pl (M.P.)

\* Correspondence: piotr.cysewski@cm.umk.pl

**Abstract:** Deep eutectic solvents (DESs) are popular green media used for various industrial, pharmaceutical, and biomedical applications. However, the possible compositions of eutectic systems are so numerous that it is impossible to study all of them experimentally. To remedy this limitation, the solubility landscape of selected active pharmaceutical ingredients (APIs) in choline chloride- and betaine-based deep eutectic solvents was explored using theoretical models based on machine learning. The available solubility data for the selected APIs, comprising a total of 8014 data points, were collected for the available neat solvents, binary solvent mixtures, and DESs. This set was augmented with new measurements for the popular sulfa drugs in dry DESs. The descriptors used in the machine learning protocol were obtained from the  $\sigma$ -profiles of the considered molecules computed within the COSMO-RS framework. A combination of six sets of descriptors and 36 regressors were tested. Taking into account both accuracy and generalization, it was concluded that the best regressor is nuSVR regressor-based predictive models trained using the relative intermolecular interactions and a twelve-step averaged simplification of the relative  $\sigma$ -profiles.

**Keywords:** solubility; green solvents; DES; machine learning; COSMO-RS;  $\sigma$ -profiles; sulfonamides



**Citation:** Cysewski, P.; Jeliński, T.; Przybyłek, M. Exploration of the Solubility Hyperspace of Selected Active Pharmaceutical Ingredients in Choline- and Betaine-Based Deep Eutectic Solvents: Machine Learning Modeling and Experimental Validation. *Molecules* **2024**, *29*, 4894. <https://doi.org/10.3390/molecules29204894>

Academic Editors: Pradip K. Bhowmik and Biplab Banerjee

Received: 27 September 2024

Revised: 13 October 2024

Accepted: 14 October 2024

Published: 16 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Active pharmaceutical ingredients (APIs) are the biologically active components of drugs that produce the intended therapeutic effects. They are the key compounds responsible for diagnosing, treating, or preventing diseases in patients [1,2]. The development and optimization of APIs are central to pharmaceutical research, as their chemical properties, such as stability, solubility, and bioavailability, significantly influence a drug's safety, efficacy, and dosage. Solubility is indeed one of the critical parameters in the development and characterization of active pharmaceutical ingredients [3–5], which affects the full lifecycle of a drug, starting with its synthesis and ending with its administration to the patient [6–8]. For an API to exert its pharmacological effect, it must first dissolve in bodily fluids, enabling its absorption into the bloodstream and subsequent distribution to target sites. Poor solubility can limit a drug's absorption, reducing its bioavailability [9,10] and potentially necessitating higher doses, which can lead to adverse effects or therapeutic failure [11–13]. Consequently, optimizing solubility is a fundamental aspect of drug formulation and delivery, ensuring that the APIs achieve the desired clinical outcomes while maintaining safety and efficacy profiles [14–17]. Following this line of thought, many techniques have been proposed to enhance the solubility of poorly dissolvable APIs [18–21]. While experimental measurements are necessary in order to obtain a full understanding of the behavior of a particular API in a selected solvent, the prediction of solubility is also extremely useful, particularly during the development of a drug [22–26]. If the number of potentially

useful solvents can be narrowed down by an initial predictive and screening stage, it is particularly valuable from an environmental perspective since it reduces the chemicals and energy needed for additional experiments. This approach fits perfectly with the concept of “green chemistry” [27,28], and it is no surprise that it was widely adopted [29–31]. The need for accurate, reliable, and relatively fast prediction methods focused the attention of researchers on neural networks and machine learning [32,33]. These techniques have already found numerous applications in the pharmaceutical domain, including solubility predictions [34–37].

The “green chemistry” approach mentioned above establishes not only the reduction of the environmental impact by limiting the number of laboratory operations, but also the introduction of more environmentally friendly solvents, the so-called “green solvents” [38,39]. These solvents, characterized by a lower toxicity and enhanced biodegradability, have quickly found widespread usage and become an alternative to the classical organic solvents [40,41]. Among the different solvents that offer environmental safety, the deep eutectic solvents (DESs) are one of the most proficient dissolution media [42–45]. These systems can be broadly defined as mixtures of two or more components, usually containing a hydrogen bond acceptor (HBA) and a hydrogen bond donor (HBD), with the melting point of the mixture being lower than the melting points of the individual constituents [46,47]. DESs combine a high solubilizing potential for many APIs with desirable properties, which includes a high potential for tuning [48–52]. The pharmaceutical industry has embraced the beneficial effects of deep eutectic solvents, which have found numerous applications in this field [53–61].

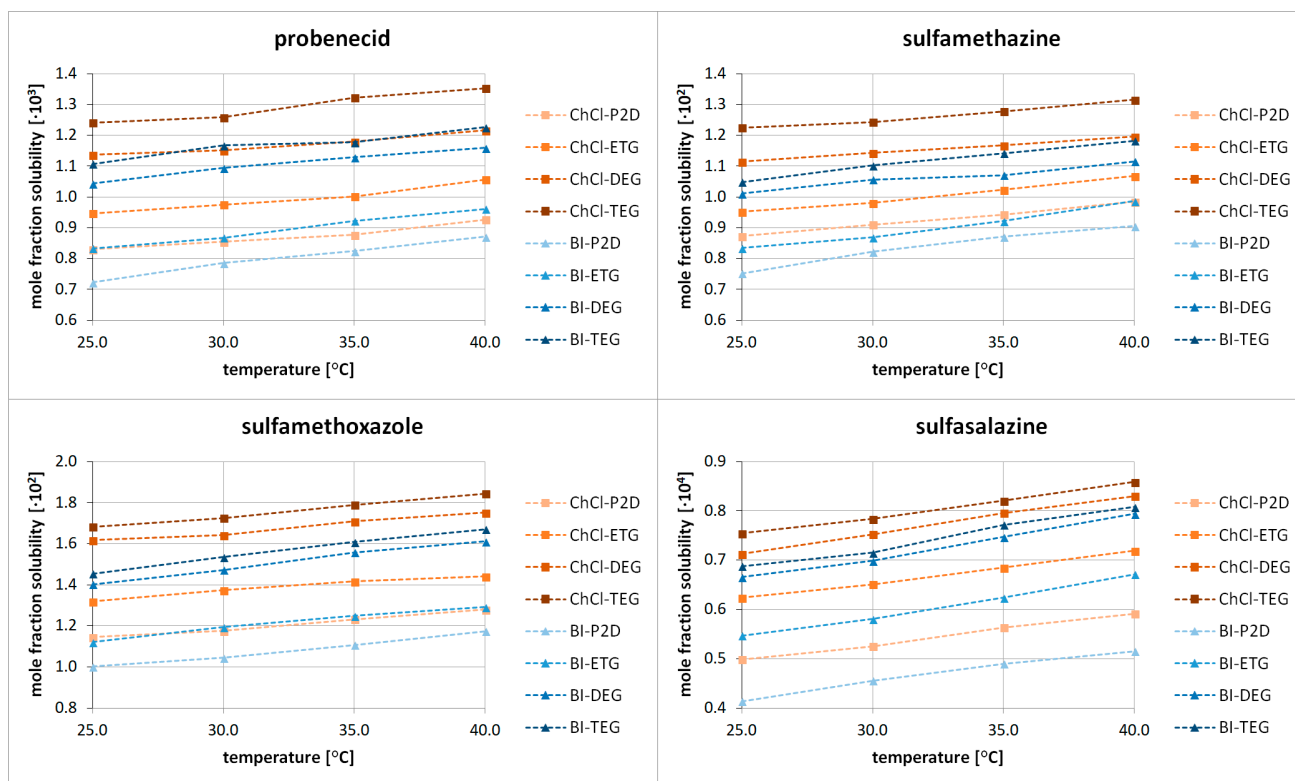
The current study is focused on the development of an effective and reliable predictive model for solubility estimation of active pharmaceutical ingredients in deep eutectic solvents based on choline chloride and betaine. For this purpose, the available solubility data space of 15 selected APIs was explored and supplemented with new experimental data. The COSMO-RS approach was employed to obtain a set of descriptors that characterize the system under consideration and were used in the machine learning protocol. This procedure resulted in the obtainment of the predictive model used for the exploration of the solubility hyperspace of the selected APIs.

## 2. Results and Discussion

### 2.1. Experimental Extension of the Solubility Dataset in DESs

The solubility dataset of various active pharmaceutical ingredients in different DES formulations is quite extended. Our research group has contributed to these efforts by measuring the DES solubility of such APIs as ibuprofen and ketoprofen [62], ferulic acid [63], curcumin [64], caffeine [65], theobromine [66], theophylline [67], dapsone [68], edaravone [69], as well as various sulfonamides, including probenecid, sulfamethazine, sulfamethoxazole, sulfasalazine, sulfacetamide, and sulfanilamide [70]. The undisputed advantage of this particular dataset is its consistency, both in terms of the applied measurement protocol and the used eutectic formulations. Nonetheless, there is always room for extending the known solubility space of APIs by conducting new measurements, and such an extension, comprising 128 new data points, is provided here.

New solubility values were obtained for four sulfonamides, namely probenecid (PC), sulfamethazine (SMZ), sulfamethoxazole (SMA), and sulfasalazine (SSZ). Eight distinct eutectic compositions were used by combining two hydrogen bond acceptors (HBAs), i.e., choline chloride (ChCl) and betaine (BI) with four hydrogen bond donors (HBDs), namely 1,2-propanediol (P2D), ethylene glycol (ETG), diethylene glycol (DEG), and triethylene glycol (TEG). The HBA:HBD molar ratio was set to 1:2, and the solubilities were measured at four different temperatures, namely 25 °C, 30 °C, 35 °C, and 40 °C. The obtained solubility values are collected in Figure 1 and in Table S1 in the Supplementary Materials.



**Figure 1.** Mole fraction solubilities of four different sulfonamides in DESs composed of choline chloride (ChCl) or betaine (BI) with 1,2-propanediol (P2D), ethylene glycol (ETG), diethylene glycol (DEG), or triethylene glycol (TEG) in a 1:2 molar ratio at various temperatures.

The general picture emerging from the obtained results shows that sulfamethoxazole is characterized by the highest solubility in the studied eutectic formulations, followed closely by sulfamethazine. The solubility of probenecid is about one order of magnitude lower, while the solubility of sulfamethoxazole has been found to be the lowest among the studied sulfonamides. This general observation is in accordance with the solubilities of these compounds in the DES systems studied earlier (see Figure 2). When it comes to the efficiency of the studied eutectic systems, it can be concluded that eutectics utilizing choline chloride have a better overall performance than those with betaine. The influence of the type of HBD is also very important from the perspective of sulfonamide solubility. In all studied cases, triethylene glycol proved to be the most efficient, followed by diethylene glycol, ethylene glycol, and 1,2-propanediol. When considering all the possible systems, a general decreasing trend of solubility can be seen, namely the following: ChCl-TEG > ChCl-DEG > BI-TEG > BI-DEG > ChCl-ETG > ChCl-P2D > BI-ETG > BI-P2D. Also, quite obviously, the temperature increase promotes the solubility of all the sulfonamides. The mole fraction solubilities at 25 °C for the best-performing ChCl-TEG system were the following:  $x_{SMA} = 0.01681$ ,  $x_{SMZ} = 0.01225$ ,  $x_{PC} = 0.00124$ , and  $x_{SSZ} = 0.000075$ . The detailed results can be found in the Supplementary Materials.

The above results should be placed in the context of the solubility values previously obtained for various APIs. Such a comparison is provided in Figure 2, which shows the collection of solubilities of the considered solutes in different DES formulations. For clarity of presentation, only the results obtained for the 1:2 molar proportion of HBA:HBD are shown, and the temperature is restricted to 25 °C.

		caffeine theobromine theophylline ferulic acid edaravone ibuprofen ketoprofen curcumin dapsona probenecid sulfacetamide sulfamethazine sulfamethoxazole sulfanilamide sulfasalazine																
		HBA	HBD	CAF	THB	THP	FA	EDA	IB	KP	CUR	DAP	PC	SCM	SMZ	SMA	SNM	SSZ
choline chloride	P2D	-	-	-	-1.4	-1.2	-0.7	-1.4	-	-2.3	-3.1	-	-2.1	-1.9	-	-4.3		
	ETG	-	-	-	-1.3	-0.8	-0.5	-1.3	-	-2.0	-3.0	-	-2.0	-1.9	-	-4.2		
	DEG	-	-	-	-1.3	-0.7	-0.5	-1.2	-	-1.8	-2.9	-	-2.0	-1.8	-	-4.1		
	TEG	-	-	-	-1.3	-0.7	-0.5	-1.2	-	-1.8	-2.9	-	-1.9	-1.8	-	-4.1		
	B3D	-	-	-	-1.4	-1.1	-0.7	-1.3	-	-2.2	-	-	-	-	-	-		
	GLY	-1.5	-2.9	-1.9	-1.3	-1.0	-0.5	-1.2	-	-2.1	-3.0	-0.8	-2.0	-1.8	-0.7	-4.3		
	FRU	-2.0	-3.1	-2.2	-	-	-	-	-	-	-3.1	-0.8	-2.0	-1.9	-0.9	-4.2		
	GLU	-1.8	-3.1	-2.2	-	-	-	-	-	-	-3.0	-0.8	-1.9	-1.9	-1.0	-4.2		
	SOR	-1.7	-3.2	-2.0	-	-	-	-	-	-	-2.8	-1.0	-1.8	-1.7	-0.9	-4.1		
	XYL	-1.7	-3.2	-2.1	-	-	-	-	-	-	-2.9	-1.0	-1.9	-1.8	-1.0	-4.2		
	SUC	-2.1	-3.4	-2.4	-	-	-	-	-	-	-3.0	-0.8	-2.2	-1.8	-0.8	-4.1		
	MAL	-2.1	-3.5	-2.5	-	-	-	-	-	-	-3.0	-1.0	-2.2	-1.9	-1.0	-4.1		
betaine	ETG	-	-	-	-1.4	-	-0.6	-1.3	-	-	-3.1	-	-2.1	-2.0	-	-4.3		
	P2D	-	-	-	-1.5	-	-0.7	-1.4	-	-	-3.1	-	-2.1	-2.0	-	-4.4		
	DEG	-	-	-	-1.4	-	-0.6	-1.3	-	-	-3.0	-	-2.0	-1.9	-	-4.2		
	TEG	-	-	-	-1.4	-	-0.5	-1.2	-	-	-3.0	-	-2.0	-1.8	-	-4.2		
	B3D	-	-	-	-1.5	-	-0.7	-1.4	-	-	-	-	-	-	-	-		
GLY	-	-	-	-1.4	-	-0.5	-1.2	-	-	-	-	-	-	-	-			

**Figure 2.** Collection of the API solubility values, expressed as the decadal logarithm of the mole fraction, in choline chloride and betaine deep eutectic solvents in ambient conditions. Newly measured values are marked with black borders. Only the 1:2 proportion of HBA and HBD was included. Colors map the span of solubility values. APIs include the following: caffeine (CAF), theobromine (THB), theophylline (THP), ferulic acid (FA), edaravone (EDA), ibuprofen (IB), ketoprofen (KP), curcumin (CUR), dapsona (DAP), probenecid (PC), sulfacetamide (SCM), sulfamethazine (SMZ), sulfamethoxazole (SMA), sulfanilamide (SNM), and sulfasalazine (SSZ). HBDs: 1,2-propanediol (P2D), ethylene glycol (ETG), diethylene glycol (DEG), triethylene glycol (TEG), 1,3-butanediol (B3D), glycerol (GLY), fructose (FRU), glucose (GLU), sorbitol (SOR), xylitol (XYL), sucrose (SUC), and maltose (MAL).

The coloring scheme allowed a direct comparison of the solubilities at room temperature. For example, SSZ has the lowest solubility among all the studied solids and augmenting the pool of data by adding the results corresponding to the new HBDs does not change this observation. Similarly, the solubility of SMZ and SMA measured for the purpose of this project also are comparable with the previously published studies. Additionally, Figure 2 documents the significant differences in the absolute solubilities among the included APIs. This shows that DES formulations can often improve the solubility of solutes relative to other solvents, but at the same, they time can be restricted by the generally low solubility of a particular API. Despite this, deep eutectic solvents can be regarded as both universal and effective solvent systems, which has been documented numerous times. For example, in the case of caffeine [65], the optimal DES composition achieved a solubility equal to 165% of the best-performing organic solvent, i.e., DMSO. Similarly, dimethyl sulfoxide, which is the most effective classical solvent for many APIs, was outperformed by theophylline and theobromine [66,67]. In the case of dapsona [68] and edaravone [69], the best organic solvents, i.e., acetone and dichloromethane, respectively, were also inferior to many DES compositions. Of course, the results are not always that impressive, as it was evidenced in the case of COX inhibitors [62]. While for ibuprofen DESs indeed outperformed the best-performing chloroform, for ketoprofen, rather surprisingly, they could not compete with methanol. Of course, one has to keep in mind that many solvents, although offering a high dissolution potential for many solutes, are not pharma-

ceutically acceptable. This emphasizes the advantages of DESs, which can be applied in the pharmaceutical industry, even if they offer a slightly smaller solubilizing efficiency. Also worth mentioning is the often dramatic increase in API solubility when comparing DES to water. The most striking example is curcumin [64], for which the most efficient eutectic was characterized by a 12-thousand-times higher solubility when compared to an aqueous solution. The environmental and health safety impact, although sometimes disputed to some extent [71–74], is also an important factor that favors eutectics. Of course, the difference in performance observed among the deep eutectic solvents is related to their composition. The results obtained for various APIs in different eutectic systems point out some general trends. It seems that such hydrogen bond donors as glycerol, triethylene glycol, and diethylene glycol can be regarded as particularly efficient and worth consideration in future studies. Also, in terms of hydrogen bond acceptors, choline chloride performed generally better than betaine. The above considerations clearly show the importance of DESs as dissolution media, as well as their potential for tuning to specific applications.

## 2.2. COSMO-RS Derived Solubility

It is commonly accepted [75–77] that the COSMO-RS (Conductor-like Screening Model for Realistic Solvation) is a reliable tool for thermodynamic properties characteristics of bulk systems. Although formally this approach is restricted to fluid systems, it can be extended to treating the solid liquid equilibria (SLE) if the experimental or estimated values of fusion data are provided. This makes the COSMO-RS very attractive for solubility computations. Unfortunately, one encounters problems with this approach, seriously limiting its applicability even for qualitative solubility guesses. There are two major obstacles, originating both from the inaccuracy of the model and also from experimental limitations. A key experimental limitation inherent to all models based on the thermodynamic principles of the dissolution process is the inability to determine the fusion properties for compounds that degrade before reaching their melting point, sublime, or undergo polymorphic or pseudopolymorphic transformations [65,78–82]. Additionally, the COSMO-RS is incapable of predicting the solubility of certain APIs in many cases, for example benzenesulfonamide [83], ferulic acid [63], or caffeine [65]. In such cases, a high concentration of the solutes in the saturated solutions prevents them from using the fast iterative method as it fails to converge, incorrectly predicting the total miscibility. The application of the alternative approach relying on the solution of the full SLE problem is quite time-consuming, and even this type of computation fails occasionally since non-physical immiscible liquid–liquid phases are predicted. Fortunately, this is not a common situation for typical organic solvents although it is encountered for the saturated systems in DESs. Hence an alternative way to determine solubility such as the application of machine learning-based models employing the COSMO-RS descriptors becomes attractive [63,83].

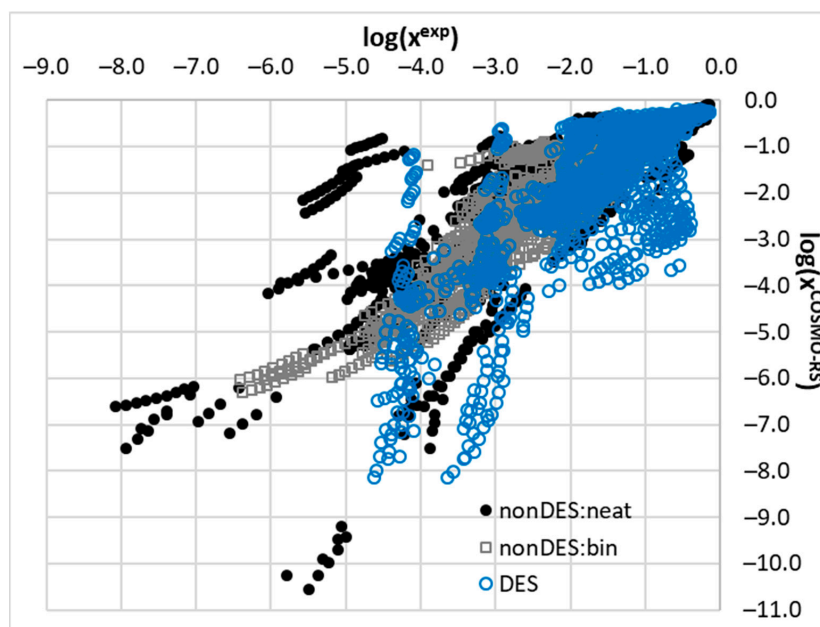
To support the above statements, a graphical representation of the relationships between the computed and measured solubility values is provided in Figure 3. It is evident that the COSMO-RS is unable to predict solubility in many cases as the majority of the computed values suffer from serious inaccuracies especially in the DESs. Among the three studied subsets, the values for the solubility of the binary mixtures were computed with the highest accuracy, which might be attributed to the quite limited range of solvents used for the measurements. This is quite understandable, since a majority of the solvent pair combinations might lead to the immiscibility of the solvents, preventing solubility measurements.

Since the main focus of this paper is on DESs, a closer inspection of this particular subset might be interesting. In Figure 4, the computed and measured solubility data are plotted for pointing out some additional limitations of COSMO-RS applicability as the source of the estimated solubility data. The plots were overlaid with open circles and open triangles as markers of the data corresponding to the new sulfonamides' measurements and the most efficient HBD, respectively. It is quite surprising that the solubility computed for sulfasalazine (SSZ) and probenecid (PC) are very inaccurate. It happened that these

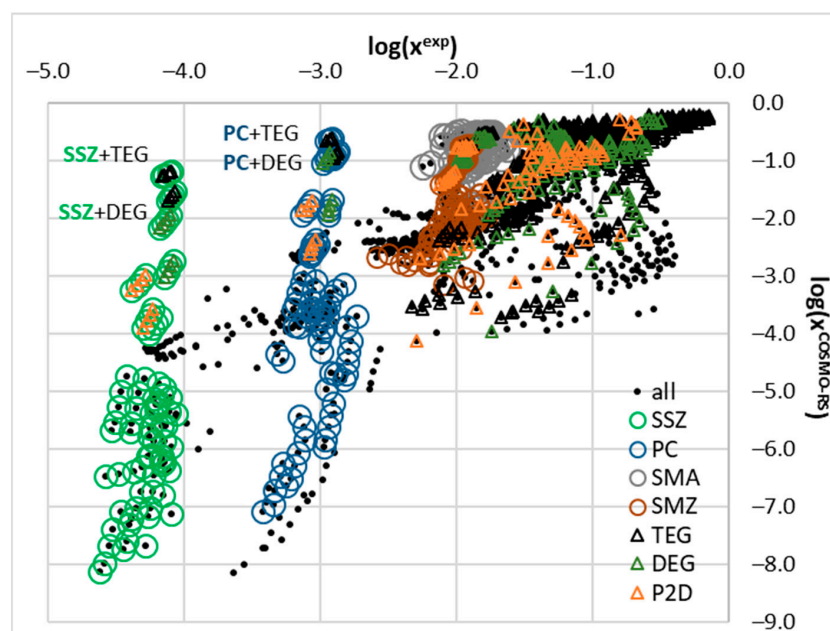




two compounds belong to a class of aromatic carboxylic acids and as such might interact with alcohols and polyalcohols such as DEG, and TEG. Interestingly, ibuprofen, ketoprofen, and ferulic acid also possess a carboxylic group but the deviation between the computed and measured values is significantly less prominent. This seems to be an interesting aspect worth subsequent study for further experimental and theoretical considerations.



**Figure 3.** Relationship of experimental solubility and values computed using COSMO-RS for all the included APIs in neat solvents, binary solvent mixtures, and all studied DES systems. The temperature relationships and concentration dependencies of saturated solutions were taken into account.



**Figure 4.** Relationship of experimental solubility and values computed using COSMO-RS for the subset of data presented in Figure 1, characterizing only APIs in DES systems. Open circles mark newly measured sulfonamides, and open triangles point out selected HBD counterparts of the studied DESs. For acronyms and their meanings, refer to Figure 1.

Being aware of the above-mentioned limitations of the COSMO-RS-derived solubility, the authors could not use such values as a direct source of the physical characteristics of the saturated systems as it is incapable of directing a proper selection of new solvents for screening purposes. However, the computed values often can be treated as a reliable source for the physicochemical characteristics of the saturated systems that enables the inclusion of the computed values, such as the valuable descriptor, for machine learning. The significant contribution of such computed solubility values to the model was already documented in our earlier works [62,63].

### 2.3. Machine Learning Solubility Model

The primary motivation for utilizing machine learning is to develop a robust model that can reliably estimate the solubility for systems not yet studied experimentally. This model is to be used as a guide for further experimental measurements and to reduce investigational efforts. By focusing on the most promising systems with a high likelihood of success, new measurements can be directed efficiently. This approach might include exploring not only different combinations of active pharmaceutical ingredients, hydrogen bond acceptors, and hydrogen bond donors but also varying proportions of deep eutectic solvent constituents and temperature dependencies. Expanding the dataset used for model development is therefore crucial for capturing the relevant factors that influence solute–solvent interactions.

In our previous studies, we examined several combinations of two different HBAs with a variety of HBDs. This led to the optimization of many practically useful solvents that meet the general requirements for pharmaceutical applications as effective dissolution media. However, despite considerable scientific effort, exploring the vast solvent hyperspace remains a challenging task, as only a small fraction of the potential combinations have been investigated. Figure 2 illustrates this limitation, showing that we covered less than 50% of the possible systems while using only a single ratio of HBA:HBD = 1:2. Clearly, exploring other solvent component ratios, even when restricting the investigation to the specified HBAs, HBDs, and water, significantly reduces the known region, leaving much more to be explored if necessary.

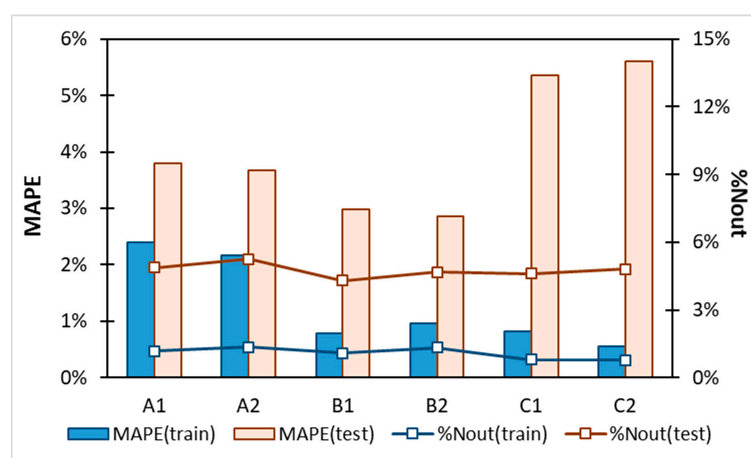
Hence, an extensive tuning of 36 regressor hyperparameters was conducted for each of the six descriptor sets to identify the models that not only best fit the experimental solubility data but also exhibited the highest predictive capability. In addition to the standard regularization parameters inherent to many regression models, a custom scoring function was employed. This function assessed the training subset's accuracy using few key metrics: the mean absolute error (MAE), as well as a penalty for the number of outliers and the number of formally not acceptable values as mentioned in the methodology Section 3.6. The final model's accuracy was evaluated using a test subset, consisting of 20% of the solubility data that were not included in the training process.

After an extensive non-linear models search, the process of regressor selection was undertaken for excluding those models that poorly represented the experimental solubility data. This resulted in the elimination of the most regressors, leaving only a few for further evaluation. The second selection criterion was an assessment of their overfitting and generalization capabilities, which was determined by comparing the accuracies of the test subsets. The models that exhibited a low mean absolute error on the training subset, but a high error on the test subset were deemed overfitted and were subsequently excluded. This group included the models generated using for example the MLPRegressor neural network, CatBoost, gradient boosting, and the XGBoost regressors. Following this analysis, the best model, demonstrating both high accuracy and strong predictive performance, was identified as the nu-Support Vector Regressor (nuSVR), a machine learning algorithm often used for regression tasks. This approach is based on the principles of Support Vector Machines (SVMs) and allows for the control of both the number of support vectors and the margin of tolerance through a parameter called “nu”, which ranges from zero to one. This flexibility makes the nuSVR effective for modeling complex, non-linear relationships,

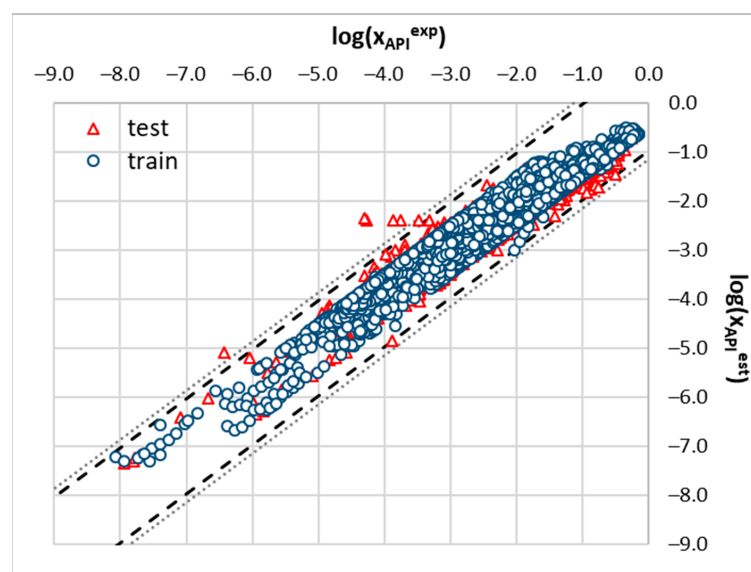


as it can balance model complexity and accuracy [84]. Recently, the nuSVR was applied successfully, predicting the solubility of aqueous solutions [85] and organic solvents [86].

Figures 5 and 6 present the final results of the nuSVR model. Instead of the mean absolute error, the mean absolute percentage error was used, as it provides a more intuitive measure of accuracy. A key observation from Figure 5 is that the choice of descriptor set has a significantly greater impact on the overall accuracy than the inclusion of the solubility data computed using the COSMO-RS. This finding is promising as it suggests that the model can be effectively applied to cases where fusion data are unavailable, thereby reducing the preparation time by omitting the solubility calculations. As illustrated in Figures 3 and 4, the computed solubility values for many of the saturated systems in this study exhibit significant inaccuracies, which does not help in model training. While the COSMO-RS has demonstrated satisfactory accuracy in previous studies, including those by our group, this is not generally the case for deep eutectic solvents. Thus, omitting the solubility values from the descriptor set is recommended based on our findings.



**Figure 5.** Comparison of nuSVR model's performance, the parameters of which were tuned on all six descriptors sets. Lines represent the percentage of outliers and bars stand for MAPE of train and test subsets.



**Figure 6.** Correlation between experimental solubility values ( $N = 8014$ ) and those computed using the NuSVR regressor trained on the B2 set of descriptors. Dotted and dashed lines represent values corresponding to three times the standard deviation computed for the whole dataset or test subset, respectively.



Another noteworthy conclusion drawn from Figure 5 is that the full  $\sigma$ -potential representation performed the worst among the tested cases. Conversely, the simplest representation using a six-step function failed to capture the essential information, making it suboptimal. The twelve-step averaging approach, which provides a balanced solute–solvent description while minimizing overfitting, is the most effective. Therefore, the descriptor set denoted as B2 is identified as the most efficient and accurate for the non-linear model training of the APIs' dissolution in DES and non-DES solvents.

### 3. Materials and Methods

#### 3.1. Materials

Four sulfonamides were used to extend the solubility dataset of the APIs in the DESs. These were as follows: probenecid (PC, CAS: 57-66-9, MW = 285.36 g/mol), sulfamethazine (SMZ, CAS: 57-68-1, MW = 278.33 g/mol), sulfamethoxazole (SMA, CAS: 723-46-6, MW = 253.28 g/mol), and sulfasalazine (SSZ, CAS: 599-79-1, MW = 398.39 g/mol). The deep eutectic solvent formulations involved in this study consisted of a hydrogen bond acceptor (HBA) and a hydrogen bond donor (HBD). Two different HBAs were used, namely choline chloride (ChCl, CAS: 67-48-1) and betaine (BI, CAS: 107-43-7). As for the HBDs, they were represented by four polyols, i.e., ethylene glycol (ETG, CAS: 107-21-1), diethylene glycol (DEG, CAS: 111-46-6), triethylene glycol (TEG, CAS: 112-27-6), and 1,2-propanediol (P2D, CAS: 57-55-6). The above compounds, both the solutes and DES constituents, were supplied by Sigma Aldrich (Saint Louis, MO, USA) and had a purity of  $\geq 99\%$ . Additionally, methanol was used as a supplementary solvent, which was delivered by Avantor Performance Materials (Gliwice, Poland) with a purity of  $\geq 99\%$ . The supplied chemicals were used without any initial procedures, apart from the choline chloride, which was dried before use.

#### 3.2. Solubility Determination

The solubility measurements were preceded by the determination of the calibration curves for the studied sulfonamides. Stock solutions of these compounds were prepared in methanol and subsequently diluted. The obtained series of solutions with decreasing concentrations were then measured spectrophotometrically using an A360 spectrophotometer from AOE Instruments (Shanghai, China) in the wavelength range of 200 nm to 500 nm. Their characteristic wavelengths were determined, and the corresponding absorbance values were plotted against the concentrations of the solutions. Three separate curves were prepared, and the obtained values were averaged. The calibration curves were validated in terms of their linearity, expressed by the determination coefficient  $R^2$ , as well as the limits of detection (LOD) and quantification (LOQ). Table 1 shows the details of the calibration curves, and based on the obtained validation values, it can be concluded that all the curves are characterized by a satisfactory linearity, as well as detection and quantification limits below the actual concentrations in the studied samples.

**Table 1.** Details of the calibration curves prepared for the sulfonamides used in the solubility dataset extension.

Compound	$\lambda_{\max}$ [nm]	Linear Regression Equation	$R^2$	LOD [mg/mL]	LOQ [mg/mL]
probenecid (PC)	246	$A = 27.628 \times C + 0.001$	0.996	0.00126	0.00378
sulfamethazine (SMZ)	269	$A = 80.729 \times C + 0.002$	0.999	0.00052	0.00157
sulfamethoxazole (SMA)	270	$A = 69.820 \times C + 0.001$	0.998	0.00067	0.00202
sulfasalazine (SSZ)	364	$A = 87.917 \times C + 0.002$	0.998	0.00042	0.00127

For determining the solubility of the studied sulfonamides in different eutectic compositions, a standard shake-flask procedure was applied [66,87–89].

The first step in this procedure involved the preparation of the deep eutectic solvents, which was performed by combining a hydrogen bond donor with a hydrogen bond acceptor. The HBA:HBD molar ratio was set to 1:2. The two HBDS, namely ChCl and BI, together with the four HBAs, namely TEG, DEG, ETG, and P2D, resulted in a total of 8 distinct eutectic compositions for each solute. The eutectics were prepared by weighing appropriate amounts of the two counterparts in glass vessels and heating them until a homogenous solution was obtained.

The samples of sulfonamides in DESs were prepared by adding an excess amount of the solute to the test tube, followed by an addition of the specific eutectic. The saturated solutions obtained in this manner were placed in an Orbital Shaker Incubator ES-20/60 from Biosan (Riga, Latvia) and heated at 25 °C, 30 °C, 35 °C, or 40 °C for 24 h with stirring at 60 rev/min. Before the measurements, the samples were filtered using a 0.22 µm pore-size PTFE syringe filter. The spectra of the filtered solutions were recorded after the dilution with methanol in the 200 nm–500 nm wavelength range with a 1 nm resolution. No solvatochromic effect was observed in the studied samples, and the analytical wavelength did not shift throughout the experiments. The absorbance values found at the characteristic wavelengths were used together with the linear regression equations to calculate the concentration of the considered solutes. Simultaneously, for the computation of the mole fraction solubility, the density of the samples was determined by weighing 1 mL of the sample on a RADWAG (Radom, Poland) AS 110 R2.PLUS analytical balance with 0.1 mg precision. Three separate samples were measured for each system, and the values were averaged.

### 3.3. Solubility Dataset

The entire solubility dataset used for machine learning, comprising  $N = 8014$  data points, includes the following fifteen active pharmaceutical ingredients: caffeine (CAF), theobromine (THB), theophylline (THP), ferulic acid (FA), edaravone (EDA), ibuprofen (IB), ketoprofen (KP), curcumin (CUR), dapsone (DAP), probenecid (PC), sulfacetamide (SCM), sulfamethazine (SMZ), sulfamethoxazole (SMA), sulfanilamide (SNM), and sulfasalazine (SSZ). The dataset is divided into three subclasses, starting with the deep eutectic solvents (DESs), which include two hydrogen bond acceptors (HBAs): choline chloride (ChCl) ( $N = 1340$ ) and betaine (BI) ( $N = 278$ ). Additionally, the newly measured data ( $N = 128$ ) were included for sulfasalazine (SSZ), sulfamerazine (SMA), sulfamethazine (SMZ), and probenecid (PC). To ensure a high degree of molecular diversity, we also collected data from the literature for each of the included APIs in the neat solvents ( $N = 2064$ ) and binary solvent mixtures ( $N = 4332$ ), when available. All data, along with references, are provided in the supporting information (see the MS Excel spreadsheet “SI.xlsx”). We believe this dataset represents the most comprehensive solubility data available for the considered APIs, given the current state of knowledge. Indeed, the dataset covers a wide range of solvent types, including highly polar protic and aprotic solvents, such as water and alcohols on one side and acetone, DMSO, and DMF on the other. The collection also includes representatives of other solvent classes, such as esters and non-polar hydrocarbons. Furthermore, halogenated solvents like dichloromethane, chloroform, and carbon tetrachloride are also present. In total, the solubility data for the 46 different solvents were included for the APIs under consideration. Interestingly, only about 33% of all the possible solute–solvent combinations are available. When accounting for temperature variations, this percentage decreases further. A coverage map, listing the measured and unavailable combinations, is provided in the Supplementary Materials (see Table S2.1). A complete documentation, along with references and solubility values, is included in the supporting materials (see the “Table.S1\_neat” spreadsheet in the “SI.xlsx” MS Excel workbook). The subset of data characterizing the binary solvent mixtures is more homogeneous, covering mixtures of water, alcohols, and some other solvents, which are listed in the Supplementary Materials (see Table S2.2). Full documentation, references, and solubility values are also included (see the “Table.S2\_bin” spreadsheet in the “SI.xlsx” workbook). Excluding the temperature and



concentration dependencies, only about 15% of the possible binary mixtures were studied for the APIs considered in this study. The third subset of the dataset includes the DESs based on choline chloride and betaine as the HBAs, mixed in various proportions with a range of HBDs. The list is available in the Supplementary Materials (see Table S2.3), and further details are available in the “SI.xlsx” file (see “Table.S3\_DES” spreadsheet). These DES systems were optimized to tailor the solubility of specific APIs. It is important to note that the addition of water to DESs is known to enhance solubility at moderate concentrations due to the nanostructuring of the HBA; however, this aspect has not been extensively studied for many API-DES systems. Formally, the coverage of possible API-HBA-HBD combinations in our dataset is about 44%, but the inclusion of varying DES component ratios and the added water significantly reduce the actual coverage. To fill the gaps in the API-DES systems, the solubility of four sulfonamides was measured in four choline chloride and betaine DESs, adding  $N = 128$  new data points to the solubility dataset.

### 3.4. COSMO-RS Solubility Computations

Solubility is routinely computed using a combination of first principles and statistical thermodynamics termed COSMO-RS (Conductor-like Screening Model for Real Solvents) [90–92]. This two-step approach leverages the macroscopic properties derived from microscopic atomistic modeling, typically employing the density functional theory (DFT) for a detailed structural diversity quantification at the molecular level. The methodology is well established [93–95], so only a brief overview is provided here. The initial step involves the generation of the relevant conformations for both the solutes and solvents. A standard protocol for an extensive conformational analysis was applied using the COSMOconf [96]. In the second step, the COSMOtherm package [97] was utilized to determine the physicochemical properties, including the solubility, intermolecular interaction contributions, and sigma potential distributions. This procedure was applied in our previous studies [62,68,69]. Each molecule was represented by up to ten low-energy conformations identified through independent conformational searches in both the gas and condensed phases. The latter is crucial for accounting for the influence of the surrounding environment within the conductor-like screening model. The outcome of this process is a set of “cosmo” and “energy” files, compatible with the latest parameter set (BP\_TZVPD\_FINE\_24.ctd). This corresponds to a two-step computational procedure: first, optimizing the molecular geometries of the most probable conformations using RI-DFT with the B88-VWN-P86 functional and def-TZVP basis set, followed by single-point energy calculations using def2-TZVPD, as implemented in Turbomole Version 7.8 (compiled 23 October 2023) [98].

The estimated solubility values correspond to the solution of the solid–liquid equilibrium under saturated conditions. This can only be achieved if the fusion data are provided as an additional input, which is essential for characterizing the activity of the pure solute under the conditions corresponding to the solubility measurements. From a general thermodynamic perspective, solubility can be computed based on the melting temperature ( $T_m$ ), the heat of fusion ( $\Delta H_{fus}$ ), and the heat capacity change upon melting ( $\Delta C_p$ ), using the following equation:

$$\ln(a^s) = -\frac{\Delta G_{fus}}{RT} = \frac{\Delta H_{fus}}{R} \cdot \left( \frac{1}{T_m} - \frac{1}{T} \right) - \frac{1}{RT} \int_{T_m}^T \Delta C_p dT + \frac{1}{RT} \int_{T_m}^T \Delta C_p dT \quad (1)$$

where  $R$  is the gas constant and  $T$  is the temperature at which the solubility is measured. For many compounds, including all fifteen APIs considered in this study, the experimental values for the melting point and heat of fusion are available [99,100]. However, the data on  $\Delta C_p$  are more difficult to measure and are unavailable for the majority of the solutes. To address this, several simplifications have been proposed [83,101–104]. In this study, we assume  $\Delta C_p \approx \Delta S_{fus} \approx \Delta H_{fus} \cdot T_m^{-1}$ . To ensure the reproducibility of the solubility calculations, all  $\Delta G_{fus}$  values have been included in the supporting materials, completing the experimental dataset (see MS Excel workbook “SI.xlsx”). The solubility values were determined using the COSMOtherm by solving the full SLE problem. While this method is

more computationally intensive than the iterative approaches, it is necessary, particularly for high solubility cases where the iterative procedure fails and inaccurately predicts the total miscibility between the solute and solvent. In contrast, the full SLE approach yields definite solubility values for every case in the dataset.

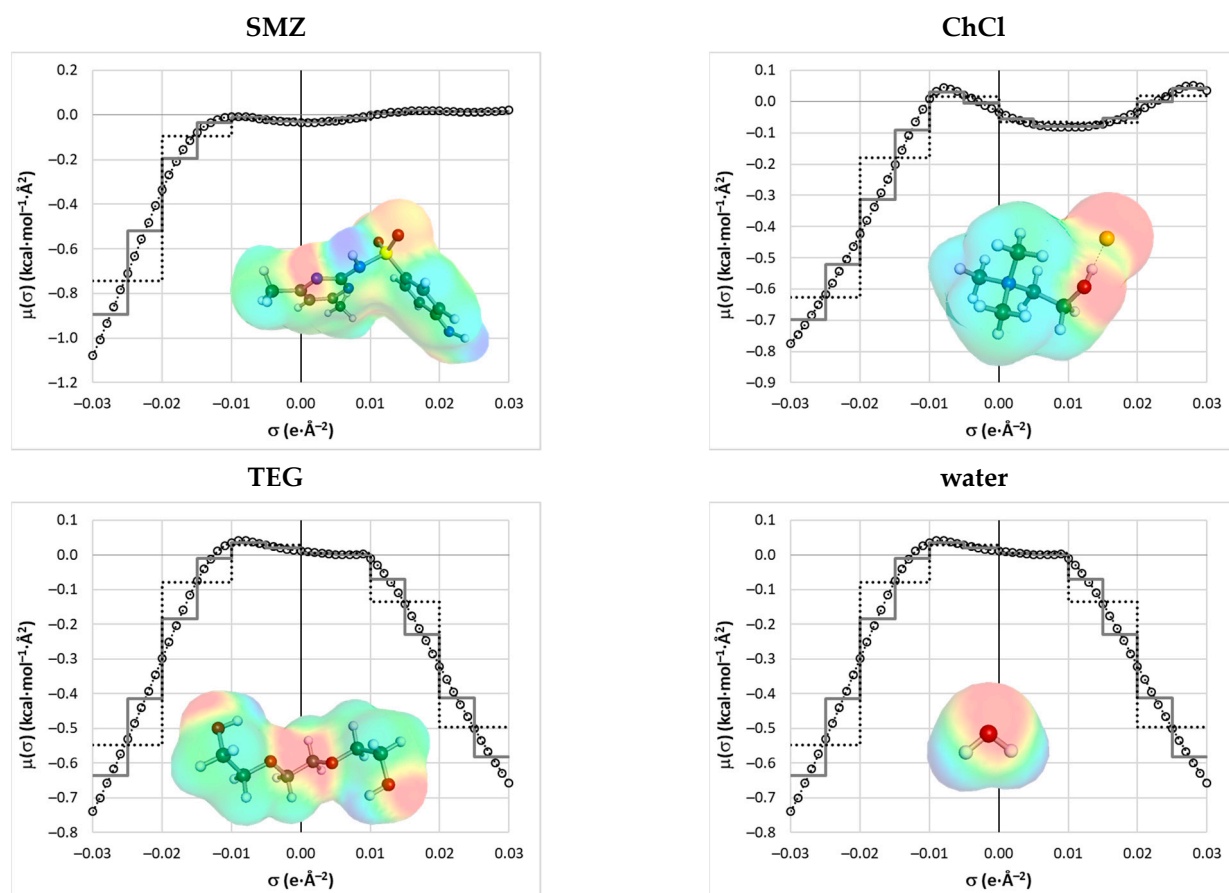
### 3.5. Molecular Descriptors

The selection of molecular descriptors is crucial for the quality of the machine learning models. The descriptors must summarize sufficient information to accurately represent the property being predicted. In the case of solubility modeling, three major requirements are essential for developing a successful model. First, the descriptors must be computable independent of the experimental data for any solute and solvent. Second, they must capture structural diversity, including isomers, rotamers, stereoisomers, and other molecular variations. Lastly, temperature dependence must be accounted for, as it plays a significant role in the equilibrium of saturated systems. These criteria exclude many popular sources of molecular descriptors, such as DRAGON [105] and PADEL [106], which rely either on SMILES strings or on 3D structural data limited to a single conformer. Fortunately, the COSMO-RS offers a comprehensive and effective way to meet these requirements. In this study, a straightforward set of descriptors was applied, leveraging the distribution of  $\sigma$ -potentials. This approach was successfully used in our previous projects [62,63], highlighting the predictive potential of this property derived from charge density distributions. For each solute and solvent, the  $\sigma$ -potential profiles were calculated for their pure, single-component state at a given temperature. The molecular descriptors used for machine learning were computed as the difference between the  $\sigma$ -potential of the pure solute and the  $\sigma$ -potential of the solvent at that specific temperature. For multicomponent solvents, the  $\sigma$ -potential was represented as a weighted value based on the solute-free mole fraction and then used to determine the relative  $\sigma$ -potential profile. Typically, the COSMOtherm program generates  $\sigma$ -potential profiles consisting of 61 points for the  $\sigma$  values ranging from  $-0.03$  to  $+0.03$  e/Å<sup>2</sup>, with a step size of 0.001. Example distributions are provided in Figure 7.

The lines with open circles represent the whole set of 61 points, defining the complete  $\sigma$ -potential of each individual component. In Figure 7, there is also a step function averaging the six subsequent values of the  $\sigma$  potentials provided. It is worth reminding that the entire range of charge densities is often split into three sub-intervals interpreted according to the encoded information. Indeed, the region of  $\sigma \in [-0.01, +0.01]$  is typically attributed to hydrophobicity (HYD), the hydrogen bond donicity (HBD) if defined by  $\sigma \in [+0.01, +0.03]$ , and the hydrogen acceptability (HBA) is addressed to  $\sigma \in [-0.03, -0.01]$ . This straightforward interpretation allows for the qualitative and quantitative characteristics of the nature of a given compound to be characterized. For example, water, being an amphiprotic agent, has non-zero contributions for the HBD and HBA regions and is insignificant for the non-polar range. Choline chloride has significantly more hydrophobic contributions compared to the other compounds presented in Figure 7, and due to the large negative volume of chloride anion exposed outwardly, it has a significant HBA range. The step function mimics the whole  $\sigma$  potential distribution with a reduced number of values. This defines the two sets of molecular descriptors, either as a 61-point full record or a simplified 12-point representation. The values presented in Figure 1 were used for the determination of the final sets of descriptors used for the machine learning protocol. For this purpose, the relative potential was defined as follows:

$$\Delta\sigma_{pot}(T) = \mu_{API}(T) - \sum_i^n x_i^* \cdot \mu_i(T) \quad (2)$$

where  $i$  iterates over all  $n$  components in the solvent mixture and is equal to 1, 2, 3, or 4 for the neat solvents, binary solvent mixtures, dry DESs, and wet DESs, respectively. Hence, the database of  $\sigma$  potentials for every compound at each temperature was prepared and used for the  $\Delta\sigma_{pot}(T)$  computations for every system in the solubility dataset.



**Figure 7.** The presentation of the individual  $\sigma$ -potentials enabling the computation of the actual values of  $\Delta\sigma_{\text{pot}}$  for SMZ + ChCl + TEG + water at  $T = 25\text{ }^{\circ}\text{C}$ . Three types of resolutions are plotted, illustrating the three sets of molecular descriptors used in the machine learning protocol, namely open circles, which denote a full distribution of 61 points (descriptors set C) and two ways of averaging. The grey solid step line represents every six points (descriptor set B), and the dotted black step line characterizes the average over twelve points (descriptor set A).

Finally, taking advantage of the solubility computations, the energetic contributions of every component were extracted from the output files for characterizing the relative values of  $\Delta\mu^{\text{mix}}$ ,  $\Delta E^{\text{tot}}$ ,  $\Delta E^{\text{HB}}$ ,  $\Delta E^{\text{misfit}}$ , and  $\Delta E^{\text{vDW}}$ . The first quantity represents the chemical potential of the mixture, and the rest characterize the contributions of the energetic components, including the relative values of the total energy, hydrogen bonding, the electrostatic contributions, and the non-polar interactions, respectively. These values were computed analogously to the formula provided in Equation (2). The training of non-linear regressors was performed independently for the four sets of descriptors, as defined in Table 2.

The three sets of descriptors A, B, and C differ significantly by the number of values used for training purposes. The comparison of model performances allows for the learning of how detailed a description of the  $\sigma$  potential should be. In the most extended case, i.e., set C, a very detailed representation is taken into account which might lead to overfitting. On the other hand, the representation of the shape of the  $\sigma$  potential by the step function might be effective since the function is generally quite smooth, and the representation of the three major regions just by the four (set B) or two (set A) averaged values might be sufficiently effective as it provides essentially the same information. It is also important to note that the motivation for excluding the values of the COSMO-RS-derived solubility is the relatively low accuracy of these estimates. This is discussed in more detail in the Section 2.2.



**Table 2.** Definition of the sets of descriptors used for the hyperparameter tuning of the considered regressors.

Model	Descriptors Set	$N_{\text{descr}}$
A1	$\Delta\sigma$ - relative potential profiles simplified by step function ( $N_{\text{descr}} = 6$ ) intermolecular interactions ( $N_{\text{descr}} = 5$ ) COSMO-RS derived solubility ( $N_{\text{descr}} = 1$ )	12
A2	similar to model A1 but without COSMO-RS derived solubility	11
B1	$\Delta\sigma$ - relative potential profiles simplified by step function ( $N_{\text{descr}} = 12$ ) intermolecular interactions ( $N_{\text{descr}} = 5$ ) COSMO-RS derived solubility ( $N_{\text{descr}} = 1$ )	18
B2	similar to model A1 but without COSMO-RS derived solubility	17
C1	$\Delta\sigma$ - relative potential profiles as full profile ( $N_{\text{descr}} = 61$ ) COSMO-RS derived solubility	62
C2	as model B1 is without COSMO-RS derived solubility	61

### 3.6. Machine Learning Protocol

The machine learning approach employed in this study follows the framework previously established in our earlier research endeavors [62,68,107]. Since the detailed methodology was thoroughly described in earlier publications, only a brief summary is presented here. The solubility prediction model was constructed using the custom Python version 3.10 [108] scripts developed for hyperparameter tuning across 36 distinct regression models. These models span a diverse array of algorithms, including linear regressors, boosting techniques, ensemble methods, nearest neighbors, neural networks, and other types of regressors. Hyperparameter tuning was performed using the Optuna version 3.2 [109], a widely used open-source Python package. The optimization process consisted of 5000 minimization trials, utilizing the tree-structured Parzen estimator (TPE) as the algorithm for sampling the hyperparameter space. To evaluate the performance of each regression model, a custom scoring function was developed, integrating multiple metrics to assess both accuracy and generalizability, as detailed in a previous study [63]. This scoring function incorporates the penalties derived from the learning curve analysis (LCA), performed using the scikit-learn library (version 1.2.2) during the parameter tuning process. It is important to highlight that the custom loss function employed for the solubility prediction models combines several evaluation metrics to ensure both accuracy and robustness, while also addressing the issue of overfitting. A key aspect of the scoring function is the integration of an LCA, which evaluates model performance across different training set sizes. By calculating the mean absolute error (MAE) through a five-fold cross-validation, the function tracks both the training and testing errors at incremental training sizes. Due to the computational demands of an LCA, a limited representation of the learning curve was adopted, using the five points between 50% and 100% of the total dataset. In comparison, our previous work employed only two points (50% and 100%). The extension of an LCA to include more data points aims to provide a more robust early indication of overfitting, albeit at the expense of an increased computational time. The custom loss function also incorporates the mean squared error (MSE) between the predicted and true solubility values, estimated for the training set, along with the penalty terms to further ensure prediction quality. These penalties address the false positive predictions (since the solubility, expressed in a logarithmic form, is expected to be negative) by penalizing any positive predictions based on their frequency. Additionally, the predictions with errors exceeding three standard deviations are classified as outliers and penalized accordingly, encouraging the model to minimize extreme deviations.

From a technical point of view, the whole set of descriptors was split into training (80%) and test (20%) subsets. The evaluation of the models' performance and their selection was performed using the unseen data by computing the root-mean-square deviations (RMSD)

and mean average percentage error (MAPE), which are defined according to the following well-known formula:

$$RMSD = \sqrt{\sum_i^n (\log(x_i^{exp}) - \log(x_i^{exp}))^2 / n} \quad (3)$$

$$MAPE = \frac{100\%}{n} \sum_i^n \left| \frac{\log(x_i^{exp}) - \log(x_i^{exp})}{\log(x_i^{exp})} \right| \quad (4)$$

In addition, the percentage of outliers (%N<sub>out</sub>) representing the percentage of the population with deviations higher than three times the standard deviation was used as the predictive coherence metric.

#### 4. Conclusions

The pharmaceutical industry faces various challenges. One of them is the need for a limitation on the energy used, and waste generated during experiments, according to the “green chemistry” framework. Another one is the constant search for new solubilizing media, which plays a pivotal role in the development of new drugs and improving the formulation of already developed drugs. These requirements can be met using the designed deep eutectic solvents. These solvent systems have found widespread use in the pharmaceutical realm; however, the number of potential DES compositions prohibits the experimental testing of all the combinations. The current study addresses this issue by means of a predictive model based on a machine learning protocol, which can limit the experiments required for selecting the optimal eutectic solvents.

The available solubility data for fifteen active pharmaceutical ingredients, namely caffeine, theobromine, theophylline, ferulic acid, edaravone, ibuprofen, ketoprofen, curcumin, dapson, probenecid, sulfacetamide, sulfamethazine, sulfamethoxazole, sulfanilamide, and sulfasalazine, were the basis for creating the predictive model. A comprehensive exploration of these data led to as many as 8014 data points for these APIs, which comprise the solubility of neat solvents (N = 2064), binary solvent mixtures (N = 4332), and DESs (N = 1618). In this set, there are 128 new data points for probenecid, sulfamethazine, sulfamethoxazole, and sulfasalazine in the various DESs containing choline chloride or betaine, obtained specifically for this study. The molecular predictors used for the purpose of the machine learning process were computed using the COSMO-RS framework. The solubility computations utilizing this method often lead to unsatisfactory results, especially in the case of eutectics, which prohibit the direct use of such values. However, the COSMO-RS can still provide meaningful and useful values in the form of  $\sigma$ -potentials.

In this study, we demonstrated that machine learning can effectively be used for predicting the solubility of new systems and also for filling the gap of lacking solubility data in partially characterized solvent hyperspace. The developed model can be used for an accurate prediction of solubility in untested systems, thereby guiding experimental efforts and optimizing resource allocation. Our extensive analysis revealed that expanding the dataset and exploring diverse solvent combinations significantly enhances the model’s predictive capability. The nu-Support Vector Regressor (nuSVR) has been found to be the most reliable model, achieving high accuracy and generalization, and the most suited for the aim of this paper. Importantly, the choice of descriptor set was found to impact predictive performance more than the inclusion of the COSMO-RS solubility data, suggesting that the models can be streamlined to improve efficiency. It is recommended to use a simplified version of the relative  $\sigma$ -potential in the form of a twelve-step function and omit the solubility values computed in the COSMO-RS. The final robust model is defined by the following parameters of the nuSVR:

- C (6.8251) controls the trade-off between maximizing the margin and minimizing the training error. A higher value of C results in a model that prioritizes fitting the training data closely, potentially at the risk of overfitting.

- Degree (8) is relevant when using polynomial kernels and determines the degree of the polynomial. A degree of eight indicates a highly flexible model capable of capturing the complex relationships in the data.
- Gamma (0.8358) defines the influence of a single training example. A higher gamma value means the model will try to fit the data more closely, as each point has a significant influence on the shape of the decision boundary.
- Max\_iter (61,378,442) sets the maximum number of iterations for the algorithm to converge. A high value ensures that the algorithm has sufficient iterations to find an optimal solution, which is especially important for complex models.
- Nu (0.4754) controls the proportion of support vectors and the margin of error, offering a balance between the number of support vectors used and the tolerance for deviations. An optimal value indicates a balanced trade-off, allowing the model to capture the underlying data patterns while controlling the margin of tolerance.

In this configuration, the model is optimized for capturing complex, non-linear relationships in the dataset while maintaining the robustness and generalization capabilities of the extended set of solubility data for DES and non-DES solvents.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules29204894/s1>. The MS Word document contains the following: Table S1: The mole fraction solubility ( $\cdot 10^4$ ) of the sulfonamides in the DESs studied in this work. The HBA:HBD molar ratio is set to 1:2. The standard deviation values are given in parentheses. Table S2.1: the solubility of 15 APIs in the neat solvents. Table S2.2: the solubility of 15 APIs in the binary solvent mixtures. Table S2.3: The solubility of 15 APIs in the DES. The SI.xlsx MS Excel file with the following spreadsheets: Table.S1\_neat, Table.S2\_bin, and Table.S3\_DES. The MS Excel includes the solubility data with references and molecular descriptors used for model development.

**Author Contributions:** Conceptualization, P.C.; methodology, P.C., T.J. and M.P.; software, P.C.; validation, P.C., T.J. and M.P.; formal analysis, P.C. and T.J.; investigation, P.C., T.J. and M.P.; resources, P.C.; data curation, P.C., M.P. and T.J.; writing—original draft preparation, P.C., T.J. and M.P.; writing—review and editing, P.C., T.J. and M.P.; visualization, P.C. and T.J.; supervision, P.C.; project administration, P.C.; funding acquisition, P.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data supporting the reported results are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Kumar, V.; Bansal, V.; Madhavan, A.; Kumar, M.; Sindhu, R.; Awasthi, M.K.; Binod, P.; Saran, S. Active pharmaceutical ingredient (API) chemicals: A critical review of current biotechnological approaches. *Bioengineered* **2022**, *13*, 4309–4327. [[CrossRef](#)]
2. Kałtny, M.; Frankowski, M. Impurities in Drug Products and Active Pharmaceutical Ingredients. *Crit. Rev. Anal. Chem.* **2017**, *47*, 187–193. [[CrossRef](#)] [[PubMed](#)]
3. Martínez, F.; Jouyban, A.; Acree, W.E. Pharmaceuticals solubility is still nowadays widely studied everywhere. *Pharm. Sci.* **2017**, *23*, 1–2. [[CrossRef](#)]
4. Savjani, K.T.; Gajjar, A.K.; Savjani, J.K. Drug Solubility: Importance and Enhancement Techniques. *ISRN Pharm.* **2012**, *2012*, 1–10. [[CrossRef](#)]
5. Coltescu, A.R.; Butnariu, M.; Sarac, I. The importance of solubility for new drug molecules. *Biomed. Pharmacol. J.* **2020**, *13*, 577–583. [[CrossRef](#)]
6. Yang, Z.; Yang, Y.; Xia, M.; Dai, W.; Zhu, B.; Mei, X. Improving the dissolution behaviors and bioavailability of abiraterone acetate via multicomponent crystal forms. *Int. J. Pharm.* **2022**, *614*, 121460. [[CrossRef](#)]
7. Kalam, M.A.; Alshamsan, A.; Alkholief, M.; Alsarra, I.A.; Ali, R.; Haq, N.; Anwer, M.K.; Shakeel, F. Solubility Measurement and Various Solubility Parameters of Glipizide in Different Neat Solvents. *ACS Omega* **2020**, *5*, 1708–1716. [[CrossRef](#)]

8. Kim, H.-S.; Kim, C.-M.; Jo, A.-N.; Kim, J.-E. Studies on Preformulation and Formulation of JIN-001 Liquisolid Tablet with Enhanced Solubility. *Pharmaceuticals* **2022**, *15*, 412. [[CrossRef](#)]
9. Khadka, P.; Ro, J.; Kim, H.; Kim, I.; Kim, J.T.; Kim, H.; Cho, J.M.; Yun, G.; Lee, J. Pharmaceutical particle technologies: An approach to improve drug solubility, dissolution and bioavailability. *Asian J. Pharm. Sci.* **2014**, *9*, 304–316. [[CrossRef](#)]
10. Müller, C.E. Prodrug Approaches for Enhancing the Bioavailability of Drugs with Low Solubility. *Chem. Biodivers.* **2009**, *6*, 2071–2083. [[CrossRef](#)]
11. Lipinski, C.A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235–249. [[CrossRef](#)] [[PubMed](#)]
12. Da Silva, F.L.O.; Marques, M.B.D.F.; Kato, K.C.; Carneiro, G. Nanonization techniques to overcome poor water-solubility with drugs. *Expert Opin. Drug Discov.* **2020**, *15*, 853–864. [[CrossRef](#)]
13. Das, B.; Baidya, A.T.K.; Mathew, A.T.; Yadav, A.K.; Kumar, R. Structural modification aimed for improving solubility of lead compounds in early phase drug discovery. *Bioorg. Med. Chem.* **2022**, *56*, 116614. [[CrossRef](#)] [[PubMed](#)]
14. Black, S.; Dang, L.; Liu, C.; Wei, H. On the measurement of solubility. *Org. Process Res. Dev.* **2013**, *17*, 486–492. [[CrossRef](#)]
15. Bergström, C.A.S.; Avdeef, A. Perspectives in solubility measurement and interpretation. *ADMET DMPK* **2019**, *7*, 88–105. [[CrossRef](#)] [[PubMed](#)]
16. Bhalani, D.V.; Nutan, B.; Kumar, A.; Singh Chandel, A.K. Bioavailability Enhancement Techniques for Poorly Aqueous Soluble Drugs and Therapeutics. *Biomedicines* **2022**, *10*, 2055. [[CrossRef](#)]
17. Manallack, D.T.; Yuriev, E.; Chalmers, D.K. The influence and manipulation of acid/base properties in drug discovery. *Drug Discov. Today Technol.* **2018**, *27*, 41–47. [[CrossRef](#)]
18. Merisko-Liversidge, E.; Liversidge, G.G. Nanosizing for oral and parenteral drug delivery: A perspective on formulating poorly-water soluble compounds using wet media milling technology. *Adv. Drug Deliv. Rev.* **2011**, *63*, 427–440. [[CrossRef](#)]
19. Brewster, M.E.; Loftsson, T. Cyclodextrins as pharmaceutical solubilizers. *Adv. Drug Deliv. Rev.* **2007**, *59*, 645–666. [[CrossRef](#)]
20. Korn, C.; Balbach, S. Compound selection for development—Is salt formation the ultimate answer? Experiences with an extended concept of the “100 mg approach”. *Eur. J. Pharm. Sci.* **2014**, *57*, 257–263. [[CrossRef](#)]
21. Seedher, N.; Kanojia, M. Co-solvent solubilization of some poorly-soluble antidiabetic drugs. *Pharm. Dev. Technol.* **2009**, *14*, 185–192. [[CrossRef](#)] [[PubMed](#)]
22. Boobier, S.; Hose, D.R.J.; Blacker, A.J.; Nguyen, B.N. Machine learning with physicochemical relationships: Solubility prediction in organic solvents and water. *Nat. Commun.* **2020**, *11*, 5753. [[CrossRef](#)] [[PubMed](#)]
23. Lovrić, M.; Pavlović, K.; Žuvela, P.; Spataru, A.; Lučić, B.; Kern, R.; Wong, M.W. Machine learning in prediction of intrinsic aqueous solubility of drug-like compounds: Generalization, complexity, or predictive ability? *J. Chemom.* **2021**, *35*, e3349. [[CrossRef](#)]
24. Hahnenkamp, I.; Graubner, G.; Gmehling, J. Measurement and prediction of solubilities of active pharmaceutical ingredients. *Int. J. Pharm.* **2010**, *388*, 73–81. [[CrossRef](#)] [[PubMed](#)]
25. Abraham, M.H.; Smith, R.E.; Luchtefeld, R.; Boorem, A.J.; Lou, R.; Acree, W.E. Prediction of solubility of drugs and other compounds in organic solvents. *J. Pharm. Sci.* **2010**, *99*, 1500–1515. [[CrossRef](#)] [[PubMed](#)]
26. Hewitt, M.; Cronin, M.T.D.; Enoch, S.J.; Madden, J.C.; Roberts, D.W.; Dearden, J.C. In silico prediction of aqueous solubility: The solubility challenge. *J. Chem. Inf. Model.* **2009**, *49*, 2572–2587. [[CrossRef](#)]
27. Lenoir, D.; Schramm, K.W.; Lalah, J.O. Green Chemistry: Some important forerunners and current issues. *Sustain. Chem. Pharm.* **2020**, *18*, 100313. [[CrossRef](#)]
28. Kopach, M.; Leahy, D.; Manley, J. The green chemistry approach to pharma manufacturing. *Innov. Pharm. Technol.* **2012**, *43*, 72–75.
29. González-Miquel, M.; Díaz, I. Green solvent screening using modeling and simulation. *Curr. Opin. Green Sustain. Chem.* **2021**, *29*, 100469. [[CrossRef](#)]
30. Derbenev, I.N.; Dowden, J.; Twycross, J.; Hirst, J.D. Software tools for green and sustainable chemistry. *Curr. Opin. Green Sustain. Chem.* **2022**, *35*, 100623. [[CrossRef](#)]
31. Sánchez-Camargo, A.d.P.; Bueno, M.; Parada-Alfonso, F.; Cifuentes, A.; Ibáñez, E. Hansen solubility parameters for selection of green extraction solvents. *TrAC Trends Anal. Chem.* **2019**, *118*, 227–237. [[CrossRef](#)]
32. Wu, Y.c.; Feng, J.w. Development and Application of Artificial Neural Network. *Wirel. Pers. Commun.* **2018**, *102*, 1645–1656. [[CrossRef](#)]
33. Sharifani, K.; Amini, M. Machine Learning and Deep Learning: A Review of Methods and Applications. *World Inf. Technol. Eng. J.* **2023**, *10*, 3897–3904.
34. Tosca, E.M.; Bartolucci, R.; Magni, P. Application of artificial neural networks to predict the intrinsic solubility of drug-like molecules. *Pharmaceutics* **2021**, *13*, 1101. [[CrossRef](#)] [[PubMed](#)]
35. Deng, T.; Jia, G. zhu Prediction of aqueous solubility of compounds based on neural network. *Mol. Phys.* **2020**, *118*, e1600754. [[CrossRef](#)]
36. Wesolowski, M.; Suchacz, B. Artificial Neural Networks: Theoretical Background and Pharmaceutical Applications: A Review. *J. AOAC Int.* **2012**, *95*, 652–668. [[CrossRef](#)] [[PubMed](#)]
37. Agatonovic-Kustrin, S.; Beresford, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* **2000**, *22*, 717–727. [[CrossRef](#)]



38. Becker, J.; Manske, C.; Randl, S. Green chemistry and sustainability metrics in the pharmaceutical manufacturing sector. *Curr. Opin. Green Sustain. Chem.* **2022**, *33*, 100562. [[CrossRef](#)]
39. Mishra, M.; Sharma, M.; Dubey, R.; Kumari, P.; Ranjan, V.; Pandey, J. Green synthesis interventions of pharmaceutical industries for sustainable development. *Curr. Res. Green Sustain. Chem.* **2021**, *4*, 100174. [[CrossRef](#)]
40. DeSimone, J.M. Practical approaches to green solvents. *Science* **2002**, *297*, 799–803. [[CrossRef](#)]
41. Häckl, K.; Kunz, W. Some aspects of green solvents. *Comptes Rendus Chim.* **2018**, *21*, 572–580. [[CrossRef](#)]
42. Santana-Mayor, Á.; Rodríguez-Ramos, R.; Herrera-Herrera, A.V.; Socas-Rodríguez, B.; Rodríguez-Delgado, M.Á. Deep eutectic solvents. The new generation of green solvents in analytical chemistry. *TrAC Trends Anal. Chem.* **2021**, *134*, 116108. [[CrossRef](#)]
43. Vanda, H.; Dai, Y.; Wilson, E.G.; Verpoorte, R.; Choi, Y.H. Green solvents from ionic liquids and deep eutectic solvents to natural deep eutectic solvents. *Comptes Rendus Chim.* **2018**, *21*, 628–638. [[CrossRef](#)]
44. Smith, E.L.; Abbott, A.P.; Ryder, K.S. Deep Eutectic Solvents (DESs) and Their Applications. *Chem. Rev.* **2014**, *114*, 11060–11082. [[CrossRef](#)] [[PubMed](#)]
45. El Achkar, T.; Greige-Gerges, H.; Fourmentin, S. Basics and properties of deep eutectic solvents: A review. *Environ. Chem. Lett.* **2021**, *19*, 3397–3408. [[CrossRef](#)]
46. Omar, K.A.; Sadeghi, R. Physicochemical properties of deep eutectic solvents: A review. *J. Mol. Liq.* **2022**, *360*, 119524. [[CrossRef](#)]
47. Paiva, A.; Craveiro, R.; Aroso, I.; Martins, M.; Reis, R.L.; Duarte, A.R.C. Natural Deep Eutectic Solvents—Solvents for the 21st Century. *ACS Sustain. Chem. Eng.* **2014**, *2*, 1063–1071. [[CrossRef](#)]
48. Espino, M.; de los Ángeles Fernández, M.; Gomez, F.J.V.; Silva, M.F. Natural designer solvents for greening analytical chemistry. *TrAC Trends Anal. Chem.* **2016**, *76*, 126–136. [[CrossRef](#)]
49. Xu, G.; Shi, M.; Zhang, P.; Tu, Z.; Hu, X.; Zhang, X.; Wu, Y. Tuning the composition of deep eutectic solvents consisting of tetrabutylammonium chloride and n-decanoic acid for adjustable separation of ethylene and ethane. *Sep. Purif. Technol.* **2022**, *298*, 121680. [[CrossRef](#)]
50. Cao, Y.; Tao, X.; Jiang, S.; Gao, N.; Sun, Z. Tuning thermodynamic properties of deep eutectic solvents for achieving highly efficient photothermal sensor. *J. Mol. Liq.* **2020**, *308*, 113163. [[CrossRef](#)]
51. Liu, Y.; Deak, N.; Wang, Z.; Yu, H.; Hamelers, L.; Jurak, E.; Deuss, P.J.; Barta, K. Tunable and functional deep eutectic solvents for lignocellulose valorization. *Nat. Commun.* **2021**, *12*, 5424. [[CrossRef](#)] [[PubMed](#)]
52. Hansen, B.B.; Spittle, S.; Chen, B.; Poe, D.; Zhang, Y.; Klein, J.M.; Horton, A.; Adhikari, L.; Zelovich, T.; Doherty, B.W.; et al. Deep Eutectic Solvents: A Review of Fundamentals and Applications. *Chem. Rev.* **2021**, *121*, 1232–1285. [[CrossRef](#)] [[PubMed](#)]
53. Bazzo, G.C.; Pezzini, B.R.; Stulzer, H.K. Eutectic mixtures as an approach to enhance solubility, dissolution rate and oral bioavailability of poorly water-soluble drugs. *Int. J. Pharm.* **2020**, *588*, 119741. [[CrossRef](#)]
54. Kapre, S.; Palakurthi, S.S.; Jain, A.; Palakurthi, S. DES-igning the future of drug delivery: A journey from fundamentals to drug delivery applications. *J. Mol. Liq.* **2024**, *400*, 124517. [[CrossRef](#)]
55. Jeliński, T.; Przybyłek, M.; Mianowana, M.; Misiak, K.; Cysewski, P. Deep Eutectic Solvents as Agents for Improving the Solubility of Edaravone: Experimental and Theoretical Considerations. *Molecules* **2024**, *29*, 1261. [[CrossRef](#)]
56. Duarte, A.R.C.; Ferreira, A.S.D.; Barreiros, S.; Cabrita, E.; Reis, R.L.; Paiva, A. A comparison between pure active pharmaceutical ingredients and therapeutic deep eutectic solvents: Solubility and permeability studies. *Eur. J. Pharm. Biopharm.* **2017**, *114*, 296–304. [[CrossRef](#)]
57. Nguyen, C.-H.; Augis, L.; Fourmentin, S.; Barratt, G.; Legrand, F.-X. Deep Eutectic Solvents for Innovative Pharmaceutical Formulations. In *Deep Eutectic Solvents for Innovative Pharmaceutical Formulations*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 56, pp. 41–102.
58. Liu, Y.; Wu, Y.; Liu, J.; Wang, W.; Yang, Q.; Yang, G. Deep eutectic solvents: Recent advances in fabrication approaches and pharmaceutical applications. *Int. J. Pharm.* **2022**, *622*, 121811. [[CrossRef](#)]
59. Emami, S.; Shayanfar, A. Deep eutectic solvents for pharmaceutical formulation and drug delivery applications. *Pharm. Dev. Technol.* **2020**, *25*, 779–796. [[CrossRef](#)]
60. Pedro, S.N.; Freire, M.G.; Freire, C.S.R.; Silvestre, A.J.D. Deep eutectic solvents comprising active pharmaceutical ingredients in the development of drug delivery systems. *Expert Opin. Drug Deliv.* **2019**, *16*, 497–506. [[CrossRef](#)]
61. Mustafa, N.R.; Spelbos, V.S.; Witkamp, G.J.; Verpoorte, R.; Choi, Y.H. Solubility and stability of some pharmaceuticals in natural deep eutectic solvents-based formulations. *Molecules* **2021**, *26*, 2645. [[CrossRef](#)]
62. Cysewski, P.; Jeliński, T.; Przybyłek, M.; Mai, A.; Kułak, J. Experimental and Machine-Learning-Assisted Design of Pharmaceutically Acceptable Deep Eutectic Solvents for the Solubility Improvement of Non-Selective COX Inhibitors Ibuprofen and Ketoprofen. *Molecules* **2024**, *29*, 2296. [[CrossRef](#)] [[PubMed](#)]
63. Jeliński, T.; Przybyłek, M.; Różalski, R.; Romanek, K.; Wielewski, D.; Cysewski, P. Tuning Ferulic Acid Solubility in Choline-Chloride- and Betaine-Based Deep Eutectic Solvents: Experimental Determination and Machine Learning Modeling. *Molecules* **2024**, *29*, 3841. [[CrossRef](#)] [[PubMed](#)]
64. Jeliński, T.; Przybyłek, M.; Cysewski, P. Natural Deep Eutectic Solvents as Agents for Improving Solubility, Stability and Delivery of Curcumin. *Pharm. Res.* **2019**, *36*, 116. [[CrossRef](#)] [[PubMed](#)]
65. Jeliński, T.; Cysewski, P. Quantification of Caffeine Interactions in Choline Chloride Natural Deep Eutectic Solvents: Solubility Measurements and COSMO-RS-DARE Interpretation. *Int. J. Mol. Sci.* **2022**, *23*, 7832. [[CrossRef](#)] [[PubMed](#)]



66. Jeliński, T.; Stasiak, D.; Kosmalski, T.; Cysewski, P. Experimental and theoretical study on theobromine solubility enhancement in binary aqueous solutions and ternary designed solvents. *Pharmaceutics* **2021**, *13*, 1118. [CrossRef]
67. Cysewski, P.; Jeliński, T.; Cymerman, P.; Przybyłek, M. Solvent screening for solubility enhancement of theophylline in neat, binary and ternary NADES solvents: New measurements and ensemble machine learning. *Int. J. Mol. Sci.* **2021**, *22*, 7347. [CrossRef]
68. Cysewski, P.; Jeliński, T.; Przybyłek, M. Experimental and Theoretical Insights into the Intermolecular Interactions in Saturated Systems of Dapsone in Conventional and Deep Eutectic Solvents. *Molecules* **2024**, *29*, 1743. [CrossRef]
69. Cysewski, P.; Jeliński, T.; Przybyłek, M. Intermolecular Interactions of Edaravone in Aqueous Solutions of Ethaline and Glyceline Inferred from Experiments and Quantum Chemistry Computations. *Molecules* **2023**, *28*, 629. [CrossRef]
70. Cysewski, P.; Jeliński, T. Optimization, thermodynamic characteristics and solubility predictions of natural deep eutectic solvents used for sulfonamide dissolution. *Int. J. Pharm.* **2019**, *570*, 118682. [CrossRef]
71. Lomba, L.; Ribate, M.P.; Zaragoza, E.; Concha, J.; Garralaga, M.P.; Errazquin, D.; García, C.B.; Giner, B. Deep Eutectic Solvents: Are They Safe? *Appl. Sci.* **2021**, *11*, 10061. [CrossRef]
72. De Moraes, P.; Gonçalves, F.; Coutinho, J.A.P.; Ventura, S.P.M. Ecotoxicity of Cholinium-Based Deep Eutectic Solvents. *ACS Sustain. Chem. Eng.* **2015**, *3*, 3398–3404. [CrossRef]
73. Macário, I.P.E.; Jesus, F.; Pereira, J.L.; Ventura, S.P.M.; Gonçalves, A.M.M.; Coutinho, J.A.P.; Gonçalves, F.J.M. Unraveling the ecotoxicity of deep eutectic solvents using the mixture toxicity theory. *Chemosphere* **2018**, *212*, 890–897. [CrossRef] [PubMed]
74. Nejrotti, S.; Antenucci, A.; Pontremoli, C.; Gontrani, L.; Barbero, N.; Carbone, M.; Bonomo, M. Critical Assessment of the Sustainability of Deep Eutectic Solvents: A Case Study on Six Choline Chloride-Based Mixtures. *ACS Omega* **2022**, *7*, 47449–47461. [CrossRef]
75. Kovács, A.; Neyts, E.C.; Cornet, I.; Wijnants, M.; Billen, P. Modeling the Physicochemical Properties of Natural Deep Eutectic Solvents. *ChemSusChem* **2020**, *13*, 3789–3804. [CrossRef]
76. Deglmann, P.; Hungenberg, K.-D.; Vale, H.M. Dependence of Copolymer Composition in Radical Polymerization on Solution Properties: A Quantitative Thermodynamic Interpretation. *Ind. Eng. Chem. Res.* **2021**, *60*, 10566–10583. [CrossRef]
77. Klamt, A.; Schwöbel, J.; Huniar, U.; Koch, L.; Terzi, S.; Gaudin, T. COSMO plex: Self-consistent simulation of self-organizing inhomogeneous systems based on COSMO-RS. *Phys. Chem. Chem. Phys.* **2019**, *21*, 9225–9238. [CrossRef]
78. Abraham, M.H.; Acree, W.E. Estimation of enthalpies of sublimation of organic, organometallic and inorganic compounds. *Fluid Phase Equilib.* **2020**, *515*, 112575. [CrossRef]
79. Jasim, F.; Talib, T. Some observations on the thermal behaviour of curcumin under air and argon atmospheres. *J. Therm. Anal.* **1992**, *38*, 2549–2552. [CrossRef]
80. Kulkarni, P.P.; Jafvert, C.T. Solubility of C60 in solvent mixtures. *Environ. Sci. Technol.* **2008**, *42*, 845–851. [CrossRef]
81. Manin, A.N.; Drozd, K.V.; Voronin, A.P.; Churakov, A.V.; Perlovich, G.L. A Combined Experimental and Theoretical Study of Nitrofurantoin Antibiotics: Crystal Structures, DFT Computations, Sublimation and Solution Thermodynamics. *Molecules* **2021**, *26*, 3444. [CrossRef]
82. Wang, Y.; Liu, Y.; Shi, P.; Du, S.; Liu, Y.; Han, D.; Sun, P.; Sun, M.; Xu, S.; Gong, J. Uncover the effect of solvent and temperature on solid-liquid equilibrium behavior of l-norvaline. *J. Mol. Liq.* **2017**, *243*, 273–284. [CrossRef]
83. Cysewski, P.; Jeliński, T.; Przybyłek, M. Finding the Right Solvent: A Novel Screening Protocol for Identifying Environmentally Friendly and Cost-Effective Options for Benzenesulfonamide. *Molecules* **2023**, *28*, 5008. [CrossRef] [PubMed]
84. Awad, M.; Khanna, R. Support Vector Regression. *Effic. Learn. Mach.* **2015**, 67–80. [CrossRef]
85. Ghanavati, M.A.; Ahmadi, S.; Rohani, S. A machine learning approach for the prediction of aqueous solubility of pharmaceuticals: A comparative model and dataset analysis. *Digit. Discov.* **2024**, *3*, 2085–2104. [CrossRef]
86. Vassileiou, A.D.; Robertson, M.N.; Wareham, B.G.; Soundaranathan, M.; Ottoboni, S.; Florence, A.J.; Hartwig, T.; Johnston, B.F. A unified ML framework for solubility prediction across organic solvents. *Digit. Discov.* **2023**, *2*, 356–367. [CrossRef]
87. Przybyłek, M.; Recki, Ł.; Mroczyńska, K.; Jeliński, T.; Cysewski, P. Experimental and theoretical solubility advantage screening of bi-component solid curcumin formulations. *J. Drug Deliv. Sci. Technol.* **2019**, *50*, 125–135. [CrossRef]
88. Jeliński, T.; Przybyłek, M.; Cysewski, P. Solubility advantage of sulfanilamide and sulfacetamide in natural deep eutectic systems: Experimental and theoretical investigations. *Drug Dev. Ind. Pharm.* **2019**, *45*, 1120–1129. [CrossRef]
89. Cysewski, P.; Jeliński, T.; Przybyłek, M. Application of COSMO-RS-DARE as a Tool for Testing Consistency of Solubility Data: Case of Coumarin in Neat Alcohols. *Molecules* **2022**, *27*, 5274. [CrossRef]
90. Klamt, A. Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235. [CrossRef]
91. Klamt, A. *From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier: Amsterdam, The Netherlands, 2005; ISBN 978-0-444-51994-8.
92. Klamt, A.; Eckert, F.; Arlt, W. COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures. *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1*, 101–122. [CrossRef]
93. Klamt, A. COSMO-RS for aqueous solvation and interfaces. *Fluid Phase Equilib.* **2016**, *407*, 152–158. [CrossRef]
94. Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 699–709. [CrossRef]
95. Klamt, A.; Eckert, F. COSMO-RS: A novel and efficient method for the a priori prediction of thermophysical data of liquids. *Fluid Phase Equilib.* **2000**, *172*, 43–72. [CrossRef]

96. Dassault Systèmes. *COSMOconf*, version 24.0.0; Dassault Systèmes; Biovia: San Diego, CA, USA, 2022.
97. Dassault Systèmes. *COSMOtherm*, version 24.0.0; Dassault Systèmes; Biovia: San Diego, CA, USA, 2022.
98. TURBOMOLE GmbH. *TURBOMOLE*, version 7.8; TURBOMOLE GmbH: Karlsruhe, Germany, 2023.
99. Acree, W.; Chickos, J.S. Phase Transition Enthalpy Measurements of Organic and Organometallic Compounds. Sublimation, Vaporization and Fusion Enthalpies From 1880 to 2015. Part 1. C 1–C 10. *J. Phys. Chem. Ref. Data* **2016**, *45*, 033101. [[CrossRef](#)]
100. Acree, W.; Chickos, J.S. Phase Transition Enthalpy Measurements of Organic and Organometallic Compounds and Ionic Liquids. Sublimation, Vaporization, and Fusion Enthalpies from 1880 to 2015. Part 2. C11–C192. *J. Phys. Chem. Ref. Data* **2017**, *46*, 013104. [[CrossRef](#)]
101. Nordström, F.L.; Rasmuson, Å.C. Determination of the activity of a molecular solute in saturated solution. *J. Chem. Thermodyn.* **2008**, *40*, 1684–1692. [[CrossRef](#)]
102. Svärd, M.; Valavi, M.; Khamar, D.; Kuhs, M.; Rasmuson, Å.C. Thermodynamic Stability Analysis of Tolbutamide Polymorphs and Solubility in Organic Solvents. *J. Pharm. Sci.* **2016**, *105*, 1901–1906. [[CrossRef](#)]
103. Svärd, M.; Hjorth, T.; Bohlin, M.; Rasmuson, Å.C. Calorimetric Properties and Solubility in Five Pure Organic Solvents of N-Methyl- d -Glucamine (Meglumine). *J. Chem. Eng. Data* **2016**, *61*, 1199–1204. [[CrossRef](#)]
104. Cysewski, P.; Jeliński, T.; Przybyłek, M.; Nowak, W.; Olczak, M. Solubility Characteristics of Acetaminophen and Phenacetin in Binary Mixtures of Aqueous Organic Solvents: Experimental and Deep Machine Learning Screening of Green Dissolution Media. *Pharmaceutics* **2022**, *14*, 2828. [[CrossRef](#)]
105. *Dragon*, version 7.0; Talete srl: Milan, Italy, 2014.
106. Yap, C.W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466–1474. [[CrossRef](#)]
107. Jeliński, T.; Przybyłek, M.; Różalski, R.; Cysewski, P. Solubility of dapsone in deep eutectic solvents: Experimental analysis, molecular insights and machine learning predictions. *Polym. Med.* **2024**, *54*, 15–25. [[CrossRef](#)] [[PubMed](#)]
108. Python Software Foundation. *Python Language Reference*, Version 3.10. Python Software Foundation: Wilmington, DE, USA. Available online: <http://www.python.org> (accessed on 12 October 2024).
109. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

